A TEST FOR THE PARAMETERS OF MULTIPLE LINEAR REGRESSION MODELS

CHRYSOGONUS C. NWAIGWE, P. I. UCHE and C. N. ONUOHA

(Received 3 November, 2003; Revision accepted 13 August, 2004).

ABSTRACT

A test for the parameters of multiple linear regression models is developed for conducting tests simultaneously on all the parameters of multiple linear regression models. The test is robust relative to the assumptions of homogeneity of variances and absence of serial correlation of the classical F-test. Under certain null and alternative hypotheses, the new test statistic is shown to have limiting central and noncentral chi-square distributions, respectively. A measure of efficiency due to Pitman is used to obtain the asymptotic efficiency of the new test relative to its classical counterpart. A numerical comparison of the two types of tests shows that the present test is slightly more efficient than the classical F-test.

KEY WORDS: test, multiple linear regression, parameters, robust, homogeneity of variances.

1. INTRODUCTION

Adichie (1967) developed a sign rank statistic for conducting tests simultaneously on intercepts and slopes in simple linear regression models. Jogdeo (1964) constructed a statistic for performing tests simultaneously on all the parameters of multiple linear regression models. Onuoha (1977) also generalized Adichie's technique by using a statistical model similar to that of Jogdeo in obtaining a sign-rank statistic for testing hypotheses about the complete set of parameters in multiple linear regression models.

The present paper develops an absolute-value test statistic for conducting tests similar to those of Jogdeo (1964) and Onuoha (1977).

In the above papers and in the present paper, the test statistic is a quadratic form consisting of component test statistics and the inverse of a covariance matrix, in the papers cited above, the component test statistics are defined in terms of the regression constants as well as the signs and/or ranks of the observations, while, in this paper, they are defined in terms of the observations and the absolute values of the regression constants.

Hence, for large observations, computations of values of the component test statistics are easier to obtain in this paper than in the above papers. Moreover, the new test statistic is robust relative to the assumptions of homogeneity of variances and absence of serial correlation of the classical F-test. This is because the proposed test statistic does not depend strictly on the variance of the observations but on the variance-covariance matrix of the component test statistics defined in (2.5).

2. THE PROPOSED TEST STATISTIC

Instead of using the signs and ranks of the observations as in Onuoha (1997), we have used the observations as well as the absolute values of the regression constants to construct the present statistic. The multiple linear regression model is of the form

\[ Y_{ni} = \beta_0 + \beta_i x_{ni} + \ldots + \beta_p x_{pi} + Z_{ni} \]

\[ Y_{ni} = \sum_{j=0}^{p} \beta_j x_{ji} + Z_{ni}; 1 \leq i \leq n < \infty, x_{0i} = 1 \text{ for all } i \]  

(2.1)

Where \( Y_{ni} \) are independent random variables with distributions.

CHRYSOGONUS C. NWAIGWE, Department of Statistics, University of Nigeria, Nsukka, Nigeria.
P. I. UCHE, Department of Statistics, University of Nigeria, Nsukka, Nigeria.
C. N. ONUOHA, Department of Statistics, University of Nigeria, Nsukka, Nigeria.
\[ p(Y_m \leq y \mid \beta_j) = F(y - \sum_{j=0}^{p} \beta_j x_{ji}) \]  \hspace{1cm} (2.2)

And where \( F \) is a distribution function, \( x_{ji} \) are known regression constants because they are assumed to be known without error. \( z_{ni} \) are independently and identically distributed random variables with zero means and unit variances and the \( \beta_j \)'s are the unknown parameters under test.

\[ (-\infty < \beta_j < \infty) \]

The problem is to test the following null and alternative hypotheses:

\[ H_0: \beta_j = 0 \text{ for all } j \hspace{1cm} (2.3) \]

\[ H_a: \beta_j = n^{-\frac{1}{2}} b_j \hspace{1cm} (2.4) \]

Where \( n \) is the sample size of the observations.

\( H_a: \beta_j = n^{-\frac{1}{2}} b_j \) implies that the \( \beta_j \)'s take values other than zero.

This is in line with Onuoha (1977). Thus, \( H_n \) is seen to tend to \( H_0 \) at the rate of \( n^{-\frac{1}{2}} \) as \( n \) increases. Let the component test statistics and the covariance matrix be defined respectively, as

\[ T_{ij} = n^{-\frac{1}{2}} \sum_{i=1}^{n} x_{ji} Y_m, 0 \leq j \leq P, x_{ji} = 1 \text{ for all } i \hspace{1cm} (2.5) \]

and

\[ \lambda_{jk} = \text{cov}(T_{ij}, T_{kl}), 0 \leq j, k \leq P \hspace{1cm} (2.6) \]

Where \( |.| \) is the absolute value symbol and \( \text{cov}(T_{ij}, T_{kl}) \) is the covariance of \( T_{ij} \) and \( T_{kl} \).

The proposed test statistic is a quadratic form consisting of the component test statistics \((T_{ij}, T_{kl}, \ldots, T_{np})\) and the inverse of the covariance matrix \( ||\lambda_{jk}|| \) and is given by

\[ Mn = (T_{ij}, T_{kl}, \ldots, T_{np}) ||\lambda_{jk}||^{-1} (T_{imulation}) = T_n ||\lambda_{jk}||^{-1} T_n \hspace{1cm} (2.7) \]

Where \( T_n = (T_{ij}, T_{kl}, \ldots, T_{np}) \)

and \( ||\lambda_{jk}||^{-1} \) is the inverse of the \((p+1) \times (p+1)\) matrix, \( ||\lambda_{jk}|| \), with elements in (2.6) while \( ||.|| \) is a matrix notation.

3. **LIMITING DISTRIBUTION OF \( M_n \) UNDER \( H_0 \)**

Let \( E_0, \text{Var}_0, \text{Cov}_0 \) and \( P_0 \), respectively, denote the expectation that are computed under \( H_0 \). From (2.1), (2.2) and the assumptions on \( Z_{ni} \), we get

\[ E(Y_m) = \sum_{j=0}^{p} \beta_j x_{ji} \hspace{1cm} (3.1) \]
Under $H_0$ in (2.3), we have

$E_0(Y_m) = 0$ \hspace{1cm} (3.2)

From (2.1) and the assumptions on $Z_m$, we also obtain

$\text{Var}_0(Y_m) = \text{Var}(Y_m) = 1$ \hspace{1cm} (3.3)

Using (3.2) in (2.5), we get $E_0(T_{ij}) = 0, 0 \leq j \leq p$ \hspace{1cm} (3.4)

Also using (2.5) and (3.3) in (2.6), we have

$\lambda_{ij} = \text{cov}_0 \left( n^{-12} \sum_{i=1}^n |X_{ij} | Y_m, n^{-12} \sum_{i=1}^n |X_{ij} | Y_m \right) = n^{-1} \sum_{i=1}^n |X_{ij}|^2$ For all $i$ \hspace{1cm} (3.5)

$\text{var}_0(T_{ij}) = \lambda_{ij} = \lambda_{ij}^2 = n^{-1} \sum_{i=1}^n |X_{ij}|^2 = n^{-1} \sum_{i=1}^n X_{ij}^2 \hspace{1cm} (3.6)$

since every $T_{ij}$ in (2.5) is the sum of independent random variables, the central limit theorem (Meyer(1973)) applies. If we let $L(T_{ij}|P) \rightarrow N(a, b^2)$ denote that the distribution law of $(T_{ij} - a)/b$ tends to the standard normal distribution under $P$, then we obtain from (3.4) and (3.6) that

$L(T_{ij}|P_0) \rightarrow N(0, \lambda_{ij}) \hspace{1cm} 0 \leq j \leq p \hspace{1cm} (3.7)$

Where $\lambda_{ij} = \lim \lambda_{ij}^2$ given in (3.6).

We have shown that, under $H_0$, the marginal distributions of the $T_{ij}$'s tend to normal distributions.

To prove the joint asymptotic normality of the $T_{ij}$'s, we use a well-known theorem of Cramer (1945). To this effect and for arbitrary constants $a_j (0 \leq j \leq p)$, we have

$T_{ij}^* = \sum_{j=0}^p a_j X_{ij} n^{-12} \sum_{i=1}^n X_{ij} | Y_m = n^{-1} \sum_{i=1}^n (\sum_{j=0}^p a_j X_{ij}) | Y_m \hspace{1cm} (3.8)$

Since $T_{ij}^*$ is the sum of independent random variables, $Y_m$, it follows from (3.2), (3.3) and central limit theorem that

$L(T_{ij}^*|P_0) \rightarrow N(0, \lambda^2$) \hspace{1cm} (3.9)

Where $\lambda^2 = \lim n^{-1} \sum_{i=1}^n (\sum_{j=0}^p a_j | X_{ij} |^2 + X_{ij}^2) = 1$ for all $i$ \hspace{1cm} (3.10)
We therefore conclude that
\[
L(T_{n0}, T_{n1}, \ldots, T_{np}) | \mathcal{P}_0 \to N_{p+1}(0, \| \lambda_n \|) \quad (3.11)
\]
\[
n \to \infty
\]

Applying the limit distribution of a continuous function of vector-valued random variables (Sverdrup (1952)) to (3.11),

We obtain,
\[
L(T_n | \mathcal{P}_0) = L(T_{n0}, T_{n1}, \ldots, T_{np}) | \mathcal{P}_0 \to L(T | \mathcal{P}_0) = N_{p+1}(0, \| \lambda_n \|) \quad (3.12)
\]
\[
n \to \infty \quad \to \infty
\]

and
\[
L(M_n | \mathcal{P}_0) = L(g(T_n) | \mathcal{P}_0) \to L(g(T) | \mathcal{P}_0) = L(T^* | \lambda_n^{-1} T) \quad (3.13)
\]
\[
n \to \infty \quad \to \infty
\]

where \( M_n = g(T_n) = T_n' | \lambda_n \| T_n \)

From a well known theorem on quadratic forms (Adichie (1987)),
\[
L(M_n | \mathcal{P}_0) \to \chi^2(u=p+1) \quad (3.14)
\]
\[
n \to \infty
\]

A direct consequence of this is that the critical function,
\[
\varphi(M_n | \mathcal{P}_0) = \begin{cases} 
1 & \text{if } M_n > \chi^2(\alpha, p+1) \\
0 & \text{if } M_n \leq \chi^2(\alpha, p+1)
\end{cases}
\]

Provides an asymptotic level \( \alpha \) of \( H_0 \), where \( \chi^2(\alpha, p+1) \) is the 100 \( (1-\alpha)\% \) point of the central chisquare distribution with \( (p+1) \) degrees of freedom

4. LIMITING DISTRIBUTION OF \( M_n \) UNDER \( H_n \)

Again, let \( E_n, \text{Var}_n, \text{Cov}_n \) and \( P_n \), respectively denote that the expectation, the variance, and the probability are computed under the alternative hypothesis, \( H_n \). From (2.4), (3.1) and (3.3), we have \( E_n(Y_n) = n^{-1/2} h_n \), and \( \text{Var}_n(Y_n) = \text{Var}_0(Y_n) = 1 \).

Where \( h_n = \sum_{j=1}^{p} b_j x_j, 1 \leq i \leq n, x_i = 1 \) for all \( i \) (4.3)

Using (4.1) in (2.5) and (4.2) in (2.6), we get
\[
E_n(T_{ny}) = n^{-1} \sum_{j=0}^{p} x_j | h_n = \mu_n \quad (4.4)
\]
and

\[ \lambda_{jk} = \text{Cov}_n(n^{-1} \sum_{i=1}^{n} X_i \mid Y_m, n^{-1} \sum_{i=1}^{n} X_i \mid Y_m) = n^{-1} \sum_{i=1}^{n} X_i \mid Y_m \mid \sum_{i=1}^{n} X_i \mid Y_m \]  \hspace{1cm} (4.5) \]

Equation (4.5) yields

\[ \text{Var}_n(T_{nj}) = \lambda_{jj} = \lambda_{jj}^2 = n^{-1} \sum_{i=1}^{n} X_i^2 = n^{-1} \sum_{i=1}^{n} \chi_i^2. \] \hspace{1cm} (4.6) \]

Using (4.4), (4.6) and the central limit theorem, we have

\[ L(T_{nj} \mid P_n) \rightarrow N(\mu_j, \lambda_{jj}), 0 \leq j \leq p, \] \hspace{1cm} (4.7) \]

\[ n \rightarrow \infty \]

Similarly,

\[ L \left( (T_{n0}, T_{n1}, \ldots, T_{np}) \mid P_n \right) \rightarrow N_{p+1} (\mu, \lambda_{jj}), \] \hspace{1cm} (4.8) \]

\[ n \rightarrow \infty \]

Where the R.H.S of (4.8) is the (P + 1)-variant normal distribution with mean vector \( \mu = \lim \mu_n \) and covariance matrix \( \| \lambda_{jk} \| = \lim \| \lambda_{nk} \| \), and where \( \mu_n \) and \( \| \lambda_{nk} \| \) have their elements defined in (4.4) and (4.5), respectively. For the limiting distribution of \( M_n \) under \( H_n \), we again use apply Sverdrup (1952) and another well-known theorem on quadratic forms (Adichie (1967)). By these, equations (3.13) and (4.8) imply that

\[ L \left( (T_n \mid P_n) \right) = L((T_n \mid P_n) \rightarrow L((T \mid P_n) = N_{p+1} (\mu, \| \lambda_{jk} \|)), \] \hspace{1cm} (4.9) \]

\[ n \rightarrow \infty \] \hspace{1cm} \[ n \rightarrow \infty \]

and

\[ L \left( (M_n \mid P_n) \right) = L((g(T_n) \mid P_n) \rightarrow L((g(T) \mid P_n) = L((T^* \| \lambda_{jk} \|^2 T)), \] \hspace{1cm} (4.10) \]

\[ n \rightarrow \infty \] \hspace{1cm} \[ n \rightarrow \infty \]

Where \( T \) has its usual meaning. The limiting distribution of \( M_n \) under \( H_n \) is completely specified by the well-known theorem on the distribution of quadratic forms cited earlier in this section (Adichie (1967)).

Hence,

\[ L \left( (M_n \mid P_n) \right) \rightarrow \chi^2 (p + 1, \triangle), \] \hspace{1cm} (4.11) \]

Where \( \triangle = \lim \triangle_n \). \hspace{1cm} (4.12) \]

5 \hspace{1cm} ASYMPTOTIC EFFICIENCY OF THE \( M_n \)-TEST

To obtain the asymptotic relative efficiency (ARE) of the \( M \)-test, with respect to its classical
counterpart the \( \hat{M}_n \)-test, we employ a measure of efficiency due to Pitman (Noether (1954)) which is defined as follows: if under the same sequence of alternatives like the one stated in (2.4), two test statistics have noncentral chi-square limit distributions with the same number of degrees of freedom, it is shown in Noether (1954) that the ARE of the two tests is the ratio of their noncentrality parameters. Hence, in order to obtain the ARE of the \( M_n \)-test with respect to the \( \hat{M}_n \)-test, we only need to derive the noncentrality parameter, \( \hat{\Delta}_n \), of the latter.

The classical test of \( H_0 \) assumes that \( F \) is the normal distribution function and uses the least squares (or the maximum likelihood) estimates, \( \hat{\beta}_{nj}, (0 \leq j \leq p) \). The \( \hat{M}_n \)-test, statistic (c.f. eg. Adiche (1967)) is

\[
\hat{M}_n = n \left( \hat{\beta}_{n0}, \hat{\beta}_{n1}, \ldots, \hat{\beta}_{np} \right) \left\| \gamma_{nj} \right\|^{-1} \left( \hat{\beta}_{n0}, \hat{\beta}_{n1}, \ldots, \hat{\beta}_{np} \right)
\]

\[
= n \beta_n \left\| \gamma_{nj} \right\|^{-1} \hat{\beta}_n, o \leq j, k \leq p,
\]

(5.1)

where \( \hat{\beta}_n = \left( \hat{\beta}_{n0}, \hat{\beta}_{n1}, \ldots, \hat{\beta}_{np} \right) \).

(5.2)

and \( \left\| \gamma_{nj} \right\|^{-1} \) is the inverse of the \((p + 1) \times (p + 1)\) matrix

\[
\left\| \gamma_{nj} \right\| = \left\| \tau_{nj} \right\|^{-1}, o \leq j, k \leq p
\]

(5.3)

and where \( \tau_{nj} = \lambda_{nj} = n^{-1} \sum_{i=1}^{n} x_{ijkl}, o \leq j, k \leq p, \chi_{ol} = 1 \) for all \( i \)

(5.4)

The estimates \( \hat{\beta}_{nj} (0 \leq j \leq p) \), being linear functions of normal random variables, are themselves normal.

By Sverdrup (1952) and the theorem on quadratic forms used in section three of this study, \( \hat{M}_n \) has a central chi-square distribution with \((p + 1)\) degrees of freedom under \( H_0 \).

Hence,

\[
L \left( \hat{M}_n | P_0 \right) \rightarrow \chi^2 (v = p + 1).
\]

(5.5)

\( n \rightarrow \infty \)

Besides, applying Adiche (1967) used in section four of this study, we have that, under any given alternatives of the form \( \beta_j = \beta_{ij} \) \((0 \leq j \leq b)\),

\( \hat{M}_n \) has a noncentral chi-square distribution with \((p + 1)\) degrees of freedom and noncentrality parameter given by (c.f. eg Lehman (1959))
\[ \hat{\Delta}_{n0} = n \tilde{\beta}_0 \parallel \gamma_{nk} \parallel^{-1} \beta_n, \]  

(5.7)

where \( \beta_0 = (\beta_{00}, \beta_{01}, \ldots, \beta_{0p}) \).  

(5.8)

It follows that under the sequence of near alternatives, \( H_n \), given in (2.4), \( \hat{\Delta}_{n0} \), becomes \( \hat{\Delta}_n \), \( (b_0, b_1, \ldots, b_p) \parallel \gamma_{njk} \parallel^{-1} (b_0, b_1, \ldots, b_p) = \tilde{b} \parallel \gamma_{nk} \parallel^{-1} \tilde{b} \)  

(5.9)

Where \( \tilde{b} = (b_0, b_1, \ldots, b_p) \).  

(5.10)

The above results show that  

\[ L \left( \hat{M}_n \mid P_n \right) \rightarrow \chi^2 (p + 1, \hat{\Delta}) \]  

(5.11)

Where \( \hat{\Delta} = \text{Lim } \hat{\Delta}_n \) is given in (5.9). Hence, the ARE of the \( M_n \) test with respect to the \( \hat{M}_n \) test, is given by  

\[ \text{ARE} \left( M_n, \hat{M}_n \right) = \frac{\Delta}{\hat{\Delta}}. \]  

(5.12)

6. **NUMERICAL COMPARISON OF THE \( M_n \) AND \( \hat{M}_n \) TESTS:**

Consider the data in table 6.1 below taken from Ronald (1996; P. 409) and used in Nwaigwe (2003)

<table>
<thead>
<tr>
<th>Table 6.1:</th>
<th>Younger’s Advertising Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \gamma_{ni} )</td>
<td>84</td>
</tr>
<tr>
<td>( X_{1i} )</td>
<td>13</td>
</tr>
<tr>
<td>( X_{2i} )</td>
<td>6</td>
</tr>
</tbody>
</table>

The model for analyzing the data is given by  

\[ Y_{ni} = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + Z_{ni}, 1 \leq i \leq 6 \]  

(6.1)

Where \( \beta_0, \beta_1 \) and \( \beta_2 \) are the unknown parameters under test. The problem is to test the following hypotheses:

\[ H_0 : \beta_j = 0, 0 \leq j \leq 2. \]  

(6.2)

\[ H_n : \beta_j = n^{-1/2} \beta_l, n = 6. \]  

(6.3)

For this comparison, we take \( b_j \) \((0 \leq j \leq 2)\) to be the least squares estimates, \( \hat{\beta}_{ni} \), of \( \beta_l \). These are given, respectively, by (see Nwaigwe (2003))

\[ b_0 = \hat{\beta}_{n0} = -41.4654, b_1 = \hat{\beta}_{n1} = 2.5444, b_2 = \hat{\beta}_{n2} = 2.6047 \]  

(6.4)
Using the data from Table 6.1 in (5.4),

\[
||\tau_{njk}|| = 1/6 \begin{pmatrix}
6 & 65 & 31 \\
65 & 733 & 341 \\
31 & 341 & 169
\end{pmatrix}
\]

(6.5)

From (5.3) and (6.5), we obtain

\[
||\gamma_{njk}||^{-1} = \begin{pmatrix}
1 & 10.8333 & 5.1667 \\
10.8333 & 122.1667 & 56.8333 \\
5.1667 & 56.8333 & 28.1667
\end{pmatrix}
\]

(6.6)

Equations (5.1), (6.4) and (6.6) give rise to \( \hat{M}_n = 24904.8696 \) (6.7)

Using the data from Table 6.1 in (2.5) and (3.5), we have

\[T_{n0} = 157.5838, \ T_{n1} = 1770.9811, \ T_{n2} = 872.8348\]

(6.8)

and

\[\lambda_{00} = 1, \ \lambda_{02} = 10.8333, \ \lambda_{02} = 5.1667\]

\[\lambda_{10} = 10.8333, \ \lambda_{11} = 122.1667, \ \lambda_{12} = 56.8333\]

(6.9)

From (6.9) and Nwaigwe (2003), we have

\[
||\lambda_{njk}||^{-1} = \begin{pmatrix}
33.3168 & -1.8154 & -2.4484 \\
-1.8154 & 0.2324 & -0.1359 \\
-2.4484 & -0.1359 & 0.7589
\end{pmatrix}
\]

(6.10)

Using (6.8) and (6.10) in (2.7), we obtain the value of \( \hat{M}_n \) under \( H_0 \) as

\[\hat{M}_n = 27452.62\]

(6.11)

But, under \( H_0 \), both \( \hat{M}_n \) and \( M_n \) have limiting central chi-square distribution with 3 degrees of freedom. Hence, from the chi-square table, we have

\[\chi^2 (0.05; 3) = 7.815 \text{ and } \chi^2 (0.01; 3) = 11.34,\]

(6.12)

We observe from (6.7), (6.11) and (6.12) that, if the two tests, \( \hat{M}_n \) and \( M_n \), were to be consistent while tending to the limit, both of them would reject \( H_0 \) at the 5% and 1% levels of significance. However, the \( M_n \) test would tend to reject \( H_0 \) more often than its parametric counterpart, the \( \hat{M}_n \) test.
Using (6.4) and (6.6) in (5.9), we have

\[ \hat{\Delta}_n = 4150.8116. \]  

(6.13)

For the noncentrality parameter, \( \Delta_n \), of the \( M_n - test \), we also use the data from Table 6.1 in (4.4) to obtain

\[ \mu_{n0} = 61.1964, \quad \mu_{n1} = 687.7049, \quad \mu_{n2} = 339.7713. \]  

(6.14)

Using (6.10) and (6.14) in (4.12), we have

\[ \Delta_n = 4163.1161 \]  

(6.15)

Also using (6.13) and (6.15) in (5.12), we obtain the ARE \( \hat{M}_n \) relative to \( \hat{M}_n \) as ARE \( (M_n, \hat{M}_n) = \)

\[ \Delta_n / \hat{\Delta}_n, = 100.2964\% \]  

(6.16)

7. **CONCLUSION**

From the foregoing, we observed that if the two tests were to be consistent while tending to the limit the present \( M_n \)-test would be slightly more efficient than the classical F-test.

8. **ACKNOWLEDGEMENT**

We are grateful to the Almighty God for this work. I also thank the referees who helped immensely to improve the quality of this work.

REFERENCES


